IJTRET "Voice of Innovative Minds"
doi.org/10.54473/IJTRET.2022.6308
International Journal of Trendy Research in Engineering and Technology
Volume 6 Issue 3 June 2022
ISSN NO 2582-0958

AUTOMATIC MULTI-DISEASES PREDICTION USING MACHINE LEARNING

[1]A. Akhila, [1]K. Hemalatha, [1]A. Navya, [1]B. Tejaswi, [1]K. Hemanth, [2]Ramesh Alladi

[1]Under Graduate Students, [2]Associate Professor , Department of Computer Science and Engineering,
ACE Engineering College, Hyderabad, Telangana, India
*Corresponding author .hemanthkeshanagari@gmail.com

ABSTRACT

Disease prediction has become one of the most difficult challenges in medicine in recent years. To eliminate the hazards connected with prediction, it is necessary to automate the process and notify the patient well in advance. The medical database is mostly made up of discrete data. As a result, making decisions using discrete data is a difficult task. Machine learning simplifies the process. The major purpose of this research is to give doctors a tool to diagnose diseases in their early stages. This model includes a user interface that allows users to anticipate ailments such as heart disease, Parkinson's disease, cancer, and diabetes. We utilized SVM and Logistic Regression for classification.

Keywords: —Heart diseases, Parkinson's diseases, Cancer diseases, Diabetes diseases.

I. INTRODUCTION

In the current period, nearly one person dies from heart disease every minute. In the field of health care, data science plays a critical role in processing massive amounts of data. Because disease prediction is a difficult undertaking, it is necessary to automate the process in order to eliminate potential hazards and to inform patients well in advance. The medical database is mostly made up of discrete data. As a result, making decisions using discrete data is a difficult task. Machine Learning, a branch of data mining, excels at handling big, well-formatted data sets.

Some people who have no idea about the diseases, sometimes simply ignore the symptoms of the disease which leads to serious conditions or death. So to improve these situations we have introduced this project which helps people to monitor their health conditions easily without going to hospitals every time. This project generally saves people time and money. So by using our project, people can decide whether they have to go to the hospital or not depending on their health conditions.

For the present work, we used the concept of machine learning to identify whether a person has that disease or not by collecting large amounts of data from the health department. We have utilized SVM and Logistic Regression for the classification of diseases. In the proposed model we are using a logistic regression model for the classification of heart diseases and Parkinson's disease. Support vector machine algorithm was used for the classification of diabetes and Cancer diseases. Logistic Regression used in the present study for cancer gave 92% accuracy.

II. LITERATURE SURVEY

Existing techniques have only a single disease testing model. For example, if a patient needs to check different diseases he needs to use different sites and models for checking the reports. In this project, we are providing a single model wherein you can check for multiple diseases. It helps doctors as well as the patient to check the report easily and it saves your time. In the present, we take into consideration four types of diseases namely heart disease, Parkinson's disease, cancer, and Diabetes. The existing methods of diagnosis of these diseases are discussed below.

1) Heart Disease is a complicated disease. The heart is the main part of the human body and according to a world health organization report, more than 17 million people die because of heart diseases every year. As Heart disease prediction contains so many calculations it takes more time for a doctor to identify. There are several machine and deep learning techniques available to predict heart and cardiovascular diseases. Ambrish et al[1] have used Logistic Regression (LR) techniques to classify and predict cardiovascular disease.

International Journal of Trendy Research in Engineering and Technology
Volume 6 Issue 3 June 2022
ISSN NO 2582-0958

2) Parkinson disease is one of the most serious diseases. There is no cure for the disease.
There is no automated system to check if a person has these diseases or not. For people affected with Parkinson's disease, the symptoms will not be noticeable and may take years to develop. Boxer Muhammad ,Pope John Paul II, and Adolf Hitler are some of the famous personalities affected by Parkinson's disease.

Sharanyaa et al[2] have tested Parkinson's data with Parametric and Non Parametric models to determine which model provides the higher classification accuracy.
.
3) Cancer disease prediction using machine learning was conducted by different researchers. Shaikh et al[3] have discussed in detail the Prediction of Cancer Disease using the Machine learning Approach.

4) Diabetes disease is most common in India and people used to check diabetics' reports frequently. Mir et al [4] in their work built a classifier model using the WEKA tool to predict diabetes disease by employing Naive Bayes, Support Vector Machine, Random Forest, and Simple CART algorithm. Charitha et al [5] have predicted Type-II Diabetes Prediction Using Machine Learning Algorithms.

Many of the existing studies focused on a specific condition. When a user wants to analyze diabetes, they must use one model, and when they want to analyze heart disease, they must use another model. This is a lengthy procedure. Also, if a user has multiple diseases but the existing system can only predict one of them, there is a potential that the death rate may rise as a result of not being able to predict the other sickness in advance.

III. PROPOSED Model

It is feasible to predict more than one disease at a time using the present model. As a result, there is no need for the user to go for multiple models in order to predict the diseases. It will save time, and it has the potential to lower mortality rates by predicting numerous diseases at a time.

IV .IMPLEMENTATION

1) **Data collection**: Data is collected from kaggale website and a few available API's.

2**) Data Cleaning**: Handling missing values and arranging the data in the required format.

3) **Train and test split**: Splitting the collected data into train and test data.

4) **Classification model**: Using required algorithms for prediction.

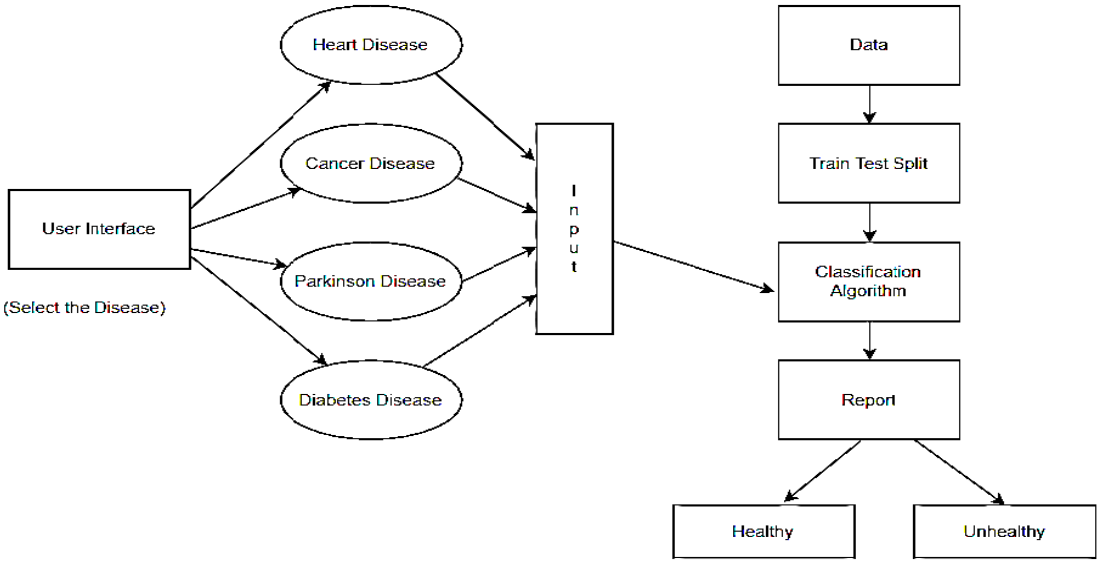

Figure 1. System Architecture

International Journal of Trendy Research in Engineering and Technology
Volume 6 Issue 3 June 2022
ISSN NO 2582-0958

Algorithms used for classification:

Logistic Regression: Under the Supervised Learning approach, one of the most common Machine Learning algorithms is logistic regression. It's a method for predicting a categorical dependent variable from a set of independent variables.

A categorical dependent variable's output is predicted using logistic regression. As a result, the result must be a discrete or categorical value. It can be Yes or No, 0 or 1, true or false, and so on, but instead of giving exact values like 0 and 1, it delivers probabilistic values that are somewhere between 0 and 1.

Except for how they are employed, Logistic Regression is very similar to Linear Regression. For regression problems, Linear Regression is employed, while for classification difficulties, Logistic Regression is used.

Instead of fitting a regression line, we fit a "S" shaped logistic function in logistic regression, which predicts two maximum values (0 or 1). The logistic function's curve reflects the probability of things like whether the cells are cancerous or not, whether a mouse is obese or not based on its weight, and so on. Because it can generate probabilities and classify new data using both continuous and discrete datasets, logistic regression is a key machine learning approach. Logistic regression can be used to categorize observations based on many forms of data and can quickly identify the most useful factors for classification. We used hyper parameters adjustment to acquire the best results for all individual datasets while testing numerous parameters.

Support vector machine (SVM):

The goal of the SVM method is to discover the best line or decision boundary for categorizing n-dimensional space into classes so that subsequent data points can be easily placed in the right category. The ideal choice boundary is known as a hyper plane.

To partition the two groups of data points, you can choose from a variety of hyper-planes. Our goal is to find the plane with the largest margin, or distance between data points from both classes. Maximizing the margin distance provides some reinforcement, making future data points easier to classify

V. RESULTS

In the system, the multiple disease prediction model used logistic regression and SVM algorithms as these gave the best accuracy accordingly. The patient needs to select the diseases that he needs to check and should provide the required data. The model will analyze the data and give a report accordingly on whether the person has a disease or not.

Table 1. ACCURACY FOR EACH DISEASE:

Disease type	Algorithm	Accuracy
Diabetes Disease	Support vector machine	**77.2**
Heart disease	**Logistic regression**	**80.4**
Parkinson' disease	**Logistic regression**	**92.1**
Cancer	**Support vector machine**	**87.1**

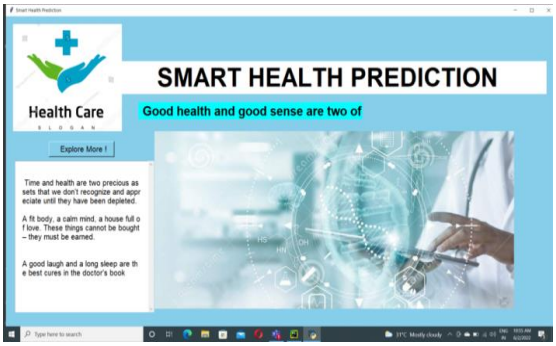

Figure 2. Screenshot of User interface

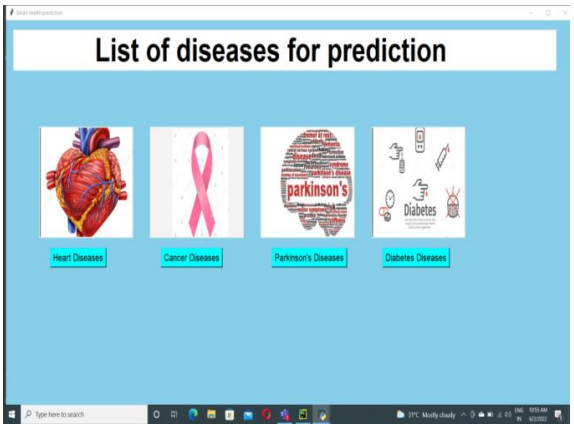

Figure 3. Screenshot showing list of diseases under study

International Journal of Trendy Research in Engineering and Technology
Volume 6 Issue 3 June 2022
ISSN NO 2582-0958

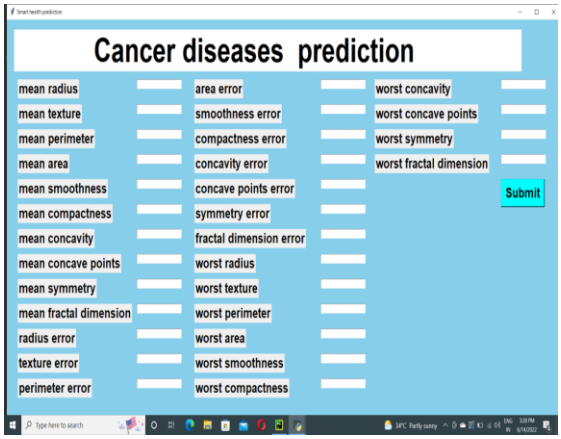

Fig.6: Screenshot showing Heart Disease data input

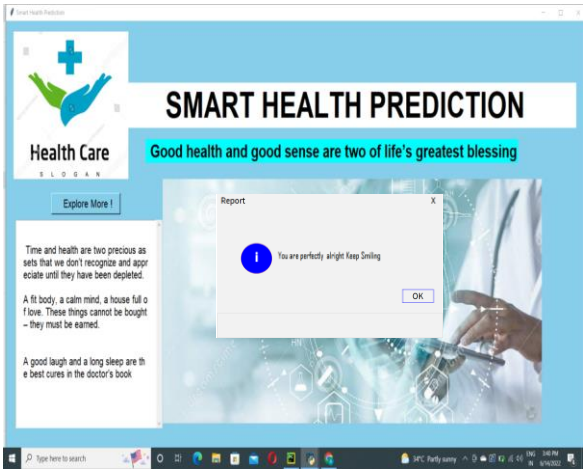

Fig.7: Screenshot showing Heart Disease Output Result

VI. CONCLUSION

The purpose of this research is to use symptoms to predict disease. The project is set up so that the system takes the user's symptoms as input and produces a disease prognosis as an output.

Disease Predictor, which tells you whether you are healthy or unwell, was built using the grails framework. Assume the patient has diabetes that has the potential to develop heart disease in the future, and then treat the patient with diabetes that prevents heart disease.

ACKNOWLEDGMENT

We are thankful to our guide Mr. Ramesh Alladi and project coordinator Mrs. Soppari Kavitha for their support and guidance. Also, we are grateful to Mr. Dr.M.V.Vijaya Saradhi, Head, Department of Computer Science and Engineering, and his support for completing the project successfully.

REFERENCES

[1] Ambrish G, Bharathi Ganesh, Anitha Ganesh, Chetana Srinivas, Dhanraj, Kiran Mensinkal, 'Logistic regression technique for prediction of cardiovascular disease', Global Transitions Proceedings,3(2022)127-130, doi.org/10.1016 /j.gltp. 2022.04.008.

[2] S.Sharanyaa,M.A.Gunavathie, Abitha.P, Sangeetha.K,, 'Classification Of Parkinson's Disease Using Logistic Regression' International Journal of Pure and Applied Mathematics 118(2018)1587-93

[3] F.J.Shaikh,D.S.Rao 'Prediction of Cancer Disease using Machine learning Approach' Materials today; Proceedings, 50(2022)40-47

[4] Mir A, Dhage SN 'Diabetes disease prediction using machine learning on big data of healthcare', IEEE, 2018 fourth international conference on computing communication control and automation (ICCUBEA) DOI:10.1109/ICCUBEA.2018.8697439

[5] C. Charitha, Amuluru Devi Chaitrasree; Penmetsa Chidananda Varma, C. Lakshmi, 'Type-II Diabetes Prediction Using Machine Learning Algorithms' 2022 International Conference on Computer Communication and Informatics (ICCCI), 2022, pp. 1-5, DOI: 10.1109/ICCCI54379.2022.9740844.